A Paradigm for Electronic Publication

David L. Steffen[1], PhD, Sinai Yarus[1], PhD, Paula E. Burch[2], PhD and David A. Wheeler[1,2], PhD
[1]Department of Cell Biology and [2]Information Technology Program,
Baylor College of Medicine, Houston, TX

The increasingly bountiful harvest of the biomedical research community is becoming an embarrassment of riches in that it results in more papers than institutions can afford to purchase or researchers have time to read. Because electronic publication (publication over the Internet) promises to solve these problems, significant efforts are underway both on the part of existing publishers[1] and by new purely electronic publishers[2] to move to an electronic publication system. Electronic publication is also capable of an immediate response to critical topics (cf. [3]). Electronic re-creation of hard copy formats is much less effective than structured databases for facilitating efficient searching. This would allow for a significant reduction in the amount of browsing required by researchers and thus the amount of time they spend monitoring the scientific literature.

As an experimental approach to determining what forms the supporting databases might take, to introduce the concept of such structured data sets to the biomedical research community, and to begin making such structured data available and useful as quickly as possible, we have developed a paradigm for electronic reviews in the form of structured databases. Here we discuss two implementations of this paradigm: the Mammary Transgene Database[4], containing information about transgenic animals where the transgenes are targeted for expression in the breast, and the Breast Cancer Information Resource[5], a repository of information linked to genes which are targets for cancer-causing mutations (e.g. proto-oncogenes and tumor suppressor genes) in breast cancer. The Mammary Transgene Database is available on the Internet, and is being actively maintained. The Breast Cancer Information Resource is less well developed, but is based closely on the previously successful Tumor Gene Database[6].

Both the Mammary Transgene Database and the Tumor Gene Database (and thus the Breast Cancer Information Resource) have been implemented as relational databases. Both are available on the World Wide Web. Both begin with MEDLINE. MeSH headings and/or text words are searched to identify appropriate references. A combination of automatic and manual methods are used to incorporate MEDLINE data into our databases, re-coded into a more searchable form. Additional information can be entered from the papers themselves. In both cases, we are developing WWW-based data entry front ends which allow for collaborative authoring. Both databases have active links to MEDLINE, Genbank/ EMBL/DDBJ/GSDB, and Entrez, with links to additional outside databases planned.

The Mammary Transgene Database is fairly rigidly organized. Typical of the kind of information contained is the level of transgene expression as RNA, protein, or activity, re-coded as high, medium, or low expression to facilitate searching. The Tumor Gene Database is more flexibly organized, with categories of information being data within the database. Both of them have been well received by authors and readers, about 10 hits a day.

In summary, we have developed an approach for electronic publication in which data are structured to facilitate searching and thus increase reader efficiency. This concept has been found to be well received by authors and readers.

Acknowledgments

References

1. Burch, P.E., MBCR's collection of on-line journals. (1996). http://condor.bcm.tmc.edu/ journals.html#full-text

2. Burch, P.E., MBCR's collection of on-line journals. (1996). http://mbcr.bcm.tmc.edu/ journals.html#e-journal

3. Yarus, S. (1996). Prion-mediated diseases in biotechnology. http://mbcr.bcm.tmc.edu/BEP/ ERMB/prion.html.

4. Yarus, S. and Steffen, D.L.,(1995). The Mammary Transgene Database. http:// mbcr.bcm.tmc.edu:80/BEP/ERMB/mtdb.html

5. Steffen, D.L. and Wheeler, D.A., (1996). http:// merlin.bcm.tmc.edu:8001/Steffen/ bcir_gene_list.html (under construction).

6. Steffen, D.L. The Tumor Gene Database. (1993). http://mbcr.bcm.tmc.edu/oncogene.html.